

Using an ongoing series of performance evaluations to drive speaker recognition technology

Alvin Martin

NIST Multimodal Information Group

International Biometric Performance Conference
March 2- 4, 2010

Outline

- Voice as a Biometric
- NIST Speaker Recognition Evaluations (SRE)
- Some SRE History
- Current NIST Programs

Voice as a Biometric

- Long recognized as a “natural” biometric
 - Most people readily recognize familiar voices
- Easy and non-invasive collection methodology
- Ubiquity of telephone system(s)
 - Supports recognition at a distance
 - Collection instruments in place
 - Use familiar to everyone
 - Routinely used for access to personal information

Performance Challenges Using Voice

- Extrinsic
 - Variability in collection devices
 - Microphone and handset type
 - Signal encoding and transmission
 - Variability in room acoustics
- Intrinsic
 - Short term
 - Emotion
 - Noise (Lombard Effect)
 - Illness
 - Speaking style
 - Long term
 - Health problems
 - Aging

NIST SREs 1996 – 2010

- Measure current state-of-the-art performance of research systems on common evaluation test sets
- Open to all interested participants worldwide
- Speech corpora collected to support evaluation
- Evaluation data distributed to all participants
- Participants submit system results to NIST
- Concluding workshops follow each evaluation
 - Participants share system descriptions
 - NIST analyzes performance results

SRE Corpora

- Successful evaluation is driven (or limited) by availability of corpora of appropriate data
- Linguistic Data Consortium (LDC) at the University of Pennsylvania has become the primary U.S. resource for speech data
 - Switchboard Corpus (~1991) laid the foundation for conversational telephone speech collection
 - Switchboard-2 (multiple phases) and Mixer Corpora have followed and been utilized in NIST SRE's

LDC Corpora Used in SRE

Name	Speakers	Conversations	Remarks
Switchboard-1	543	~2400	Groundbreaking original
Switchboard-2 Phase 1	657	3638	Most speaker from mid-Atlantic states
Switchboard-2 Phase 2	679	4472	Most speaker from Midwest
Switchboard Cellular Part 1	254	1309	Mainly GSM
Switchboard Cellular Part 2	419	2020	Mix of cellular types
Mixer 1/2	> 2224	13,769	Some multi-lingual speakers
Mixer 3	1867	19,951	Some multi-lingual speakers
Mixer 4	135	246	Includes room mic recordings
Mixer 5	~300	~1800 interviews	Mixer 3 speakers
Greybeard	171	4682	New and old calls

SRE Tasks

- Speaker Detection
 - Key task in NIST evaluations
 - Given a target speaker, determine if target is present in a test segment
 - Requires a decision ('T' or 'F') and a score – higher score indicates more probable 'T'
 - Score may be used as decision threshold to define the range of possible operating points
- Speaker Tracking
 - Determine where in the speech signal each speaker is speaking
 - Included in earlier SREs and in other NIST evaluations
- Speaker Identification
 - Select speaker out of gallery of N possibilities (open or closed set)
 - Not addressed by NIST evaluations

SRE Performance Measure (Cost Function)

- Two types of error:
 - *Miss*
 - 'F' decision when target is present (*target trial*)
 - Cost of each is C_{Miss}
 - *False Alarm*
 - 'T' decision when target not present (*non-target or impostor trial*)
 - Cost of each is $C_{\text{FalseAlarm}}$
- System calibration depends on the prior probability of a target trial P_{Target}
- NIST uses a weighted combination of the two error rates as its cost function C_{Det} (primary metric):

$$C_{\text{Det}} = C_{\text{Miss}} \times P_{\text{Miss|Target}} \times P_{\text{Target}} \\ + C_{\text{FalseAlarm}} \times P_{\text{FalseAlarm|NonTarget}} \times (1 - P_{\text{Target}})$$

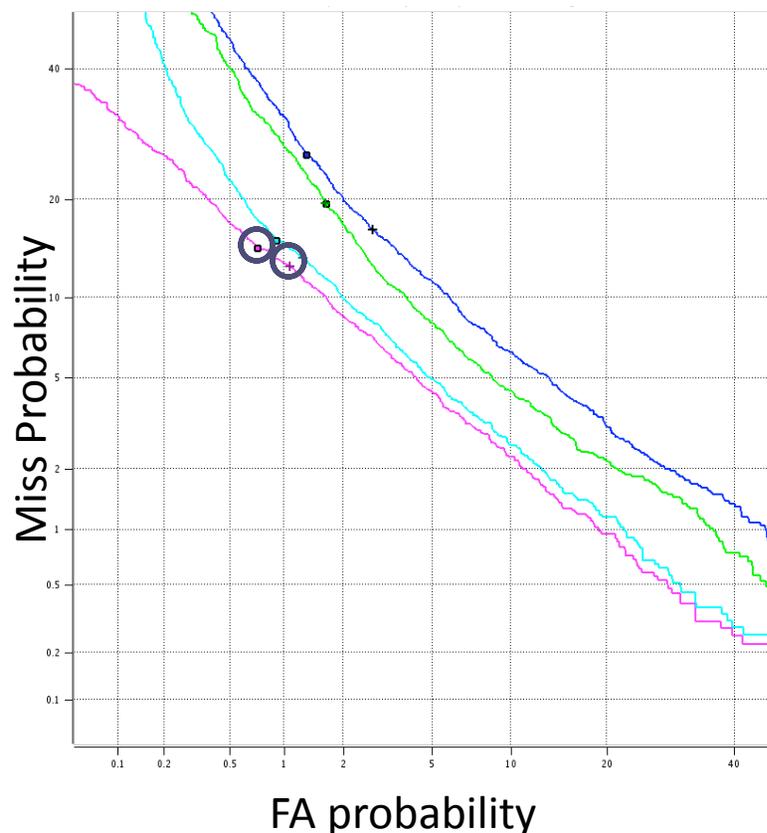
C_{Miss}	$C_{\text{FalseAlarm}}$	P_{Target}
10	1	0.01

SRE Cost Function Parameters

- NIST's C_{Det} has used a 10:1 weighting of FAs over misses
- Researchers often prefer other measures
 - Equal weighting (average of miss and FA rates)
 - Equal error rate point (ignoring calibration)
- Calibration matters, and real applications necessitate minimizing one error rate at expense of other
 - Equal error point not of primary interest for most real applications
- SRE10 will experiment with a 1000:1 ratio
 - More representative of application needs

Detection Error Tradeoff (DET) Curves

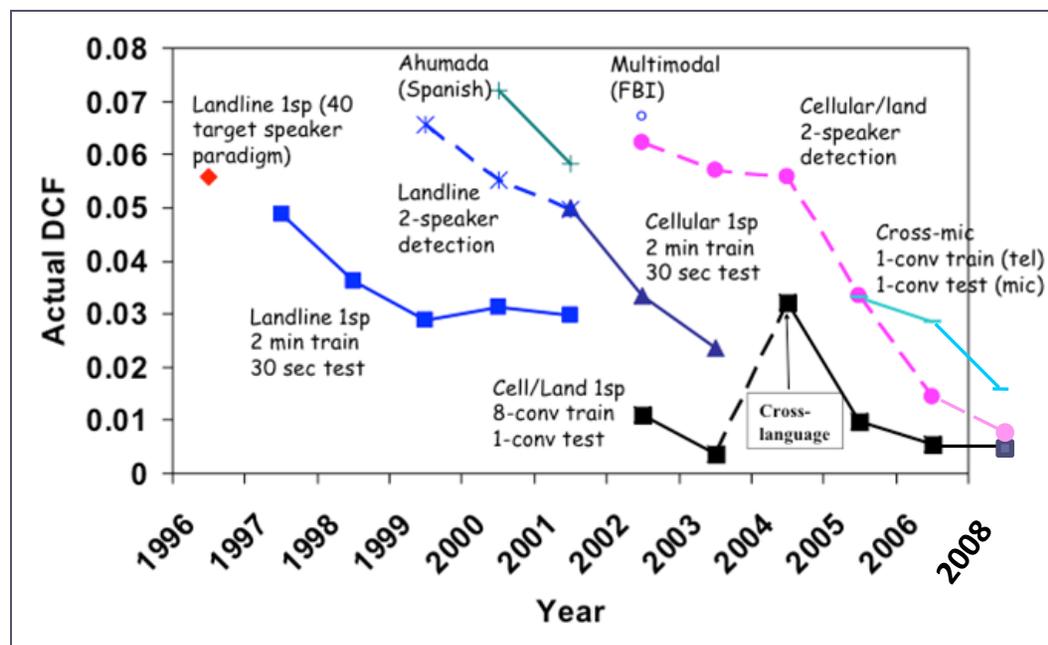
- Show range of possible operating points as decision threshold is varied
- Plot on normal deviate scale
- Actual (+) and minimum cost (o) operating point are marked
 - Distance between two points reflects threshold calibration



SRE Performance History

Best System C_{Det} (DCF) for similar test conditions

- Test conditions, and the phone system, have changed over SRE history
- Chart attempts to make apples-to-apples comparisons over multiple years of SRE
- Trend is improving performance over time



Thanks to Doug Reynolds of MIT Lincoln Lab for providing the plot

SRE Participating Sites

- Increasing participation over time
 - 1996 – 2001: 8-12
 - 2002 – 2005: 20-24
 - 2006: 36
 - 2008: 46
 - 2010: 53 (to date)
- Bulk of participants were from U.S. in earlier years, later from Europe, now from Far East
 - Includes participants from Australia, South Africa, the Middle East, and Latin America

Current NIST Speaker Programs

- SRE10
- IARPA BEST Program
- Investigatory Voice Biometrics

SRE10

- Taking place spring 2010
 - Registration closes March 1
 - Workshop June 24-25 in Brno, the Czech Republic, in conjunction with Odyssey International Workshop
- Includes conversational telephone and interview speech recorded over multiple room microphones
- New Conditions tested
 - Speakers with training and test speech recorded years apart (Greybeard Corpus)
 - Vocal effort
 - Human Assisted Speaker Recognition (HASR)

IARPA BEST Program

Biometrics Exploitation Science and Technology

- Seeks to drive research progress on face, ocular, and voice biometric technology
 - Advance the ability to achieve high-confidence match performance, despite features derived from non-ideal data
 - Significantly relax the constraints currently required to acquire high fidelity biometric signatures

IARPA BEST Program (cont'd)

Biometrics Exploitation Science and Technology

- Program kicked off December 2009
- Phase I to last two years
- Three performance teams selected for speaker recognition effort
- Evaluation at end of Phase I
 - Coordinated by NIST
 - Open to outside participants

Investigatory Voice Biometrics

- Project begun in 2009 with FBI support
- Initial workshop held at NIST in March 2009
- Will produce roadmap document to support future collection efforts of government agencies addressed through four committees
 - Use Case
 - Collection Standards
 - Interoperability
 - Science and Technology